Detection of a novel minisatellite-specific DNA-binding protein

Andrew Collick and Alec J. Jeffreys
Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Received September 7, 1989; Revised and Accepted January 10, 1990

ABSTRACT

We describe the detection of a ubiquitous DNA-binding protein which appears to interact specifically with tandem-repeated minisatellites. The murine 40kd protein, which we term Msbp-1, was found to be present in all mouse tissues tested. This protein was bound specifically and with high affinity by double-stranded DNA containing a repeat sequence related to the minisatellite 'core' sequence, and binding required the presence of multiple repeat units. Corresponding minisatellite-specific DNA-binding proteins could also be detected in species ranging from Drosophila to man. This analysis represents the first direct evidence that minisatellites can function as a specific recognition signal for an endogenous DNA-binding protein.

INTRODUCTION

Minisatellites have been found in the genomes of all vertebrate species so far analysed (1). These tandem-repetitive DNA sequences frequently show extensive allelic variation in the number of repeat units, and can provide highly informative genetic markers ideal for linkage analysis (2), individual identification (3,4), and the determination of family relationships (5). While hypervariability at minisatellites is the direct result of a high germline mutation rate to new length alleles at these loci (6), the nature of the mutation process remains obscure. The presence of a common 'core' sequence GGAGGTGGGC-AGGARG present, with minor variation, in each repeat unit of a subset of human minisatellites (7) suggests that the core sequence may be directly involved in generating these hypervariable loci, perhaps by serving as the recognition sequence for a DNA-binding protein involved in the mutagen process. Similarity between the minisatellite core sequence and the chi sequence, a recombination signal in p. coli, has led to suggestions that the core sequence may act as a recombination signal which promotes unequal crossing over at minisatellites to generate new length alleles (1,7). This suggestion is supported by the discovery of a core-related sequence at or near a meiotic recombination hotspot in the mouse MHC complex (8), by the preferential localization of human minisatellites near the ends of chromosomes in regions believed to be proficient in meiotic recombination (9), and by the preferential in situ hybridization of repeated core probes to chiasmata in human meiotic chromosomes (10). However, there is no evidence for exchange of flanking markers associated with minisatellite length change mutation events, nor for the generation of new mutant interallelic recombinant alleles, as predicted for unequal exchange between homologous chromosomes (11,12,28). Furthermore, somatic mutation at minisatellites also occurs, and has been detected in human cell lines, tumours, and blood (13,28), as well as early in mouse development (14). These results suggest that the core sequence, if it serves as a recombination signal, is not implicated in meiotic recombination, but may instead be promoting unequal exchanges preferentially between sister chromatids.

An alternative approach to studying the molecular processes which generate variability at minisatellites is to search for the putative protein(s) which interact specifically with the minisatellite core sequence. In this paper we describe the detection of such a minisatellite-specific DNA-binding protein.

MATERIALS AND METHODS

Oligonucleotide probes and southwestern blot analysis

Oligonucleotides were synthesised on an Applied Biosystems 380B DNA synthesiser using reagents supplied by Cruachem. Full length oligonucleotides were purified by electrophoresis on an 8M urea, 12% polyacrylamide sequencing gel. 32P-labelled double-stranded oligonucleotides (specific activity > 5 x 106 cpm/pmol DNA) were prepared by primer extension from the M13 universal priming site present in each oligonucleotide, using Klenow fragment and alpha-32P-dCTP (15). 50—100μg samples of nuclear proteins prepared by the method of Dignam et al (1983)(16) were separated by SDS-PAGE on 12% polyacrylamide gels (17). Gels were then soaked in 4M urea/TNE-150 (10mM Tris-HCl(pH 7.6), 150mM NaCl, 2mM EDTA) for 30 minutes at room temperature. Proteins were blotted by diffusion onto Hybond-C nitrocellulose (18), with transfer at room temperature in TNE-150 for 36—48h. Nitrocellulose strips were washed in 20ml of binding buffer: 1X Denhardt's (19), 150mM NaCl, 10mM Tris-HCl(pH 7.6), 0.1mM EDTA, 20μg/ml yeast RNA (Sigma) and 40—50μg/ml sonicated, double-stranded E. coli genomic competitor DNA (strain Y1090,[26]), for 1h at 37°C. 10ng of 32P-labelled DNA was added to the buffer, and complexes allowed to form for 1—2h at 37°C. For competition experiments, the time of binding was increased to 16h. Filters were then washed three times in TNE-150 (total volume 1 litre) at 37°C for 30 min, blotted dry, and autoradiographed at —80°C for 1—2 days, using Fuji RX X-ray film and an intensifier screen. When appropriate, densitometry was performed using an LKB...
Figure 1. Detection of minisatellite-specific DNA-binding proteins by southwestern analysis of nuclear extracts. A. Oligonucleotides used to detect binding activity. B5 contains five repeats of the core sequence (uppercase) and an M13 universal priming site (underlined) used to prepare double-stranded probe. The control oligonucleotides, R1 and R2, contain randomized versions of the minisatellite sequence present in B5. B. southwestern analysis of nuclear extracts from mouse brain (Mus musculus, Mu), rabbit brain (Oryctolagus cuniculus, Or), Xenopus laevis lung (Xe) and whole adult Drosophila melanogaster (Dr), using double-stranded probes B5, R1 and R2. C. Minisatellite binding proteins in nuclear extracts of mouse brain (Br), heart (He), kidney (Ki), liver (Li) and testis (Te) detected by oligonucleotide B5.

Ultrascan densitometer. on autoradiographs taken on preflashed film (20).

Generation of synthetic minisatellites
Minisatellite B5 was prepared from complementary overlapping minisatellite core oligonucleotides 5'GTGGGCAGGAAG 3' and 3'CCGTCCCTCCAC 5' by annealing 1μg of each kinased oligonucleotide in 50mM Tris-HCl (pH 7.6), 10mM MgCl2 at 37°C for 0.5h, followed by ligation with 10 units of T4 DNA ligase, plus 1mM ATP, at 37°C for 0.5h, to form head-to-tail
and pancreas; not shown). Variable levels of a second minisatellite-specific binding protein of 36kd were present in cells (40kd), rat testis (40kd), dog testis (40kd) and trout liver (39kd) in nuclear extracts of human EJ bladder carcinoma (40kd) and human B lymphoid cell line (40kd) equivalent minisatellite-specific DNA-binding protein has also been detected in nuclear extracts of human EJ bladder carcinoma (40kd) and human B lymphoid cell line (40kd) as well as in other tissues and species.

For studies on the affinity of DNA-binding by Msbp-1, competitor double-stranded oligonucleotides were prepared using the polymerase chain reaction, upon starting DNA molecules B, R, or R. The PCR reaction mix was as described (21), with the following pairs of oligonucleotide primers being used: for B, 5'TGTAAGAACGGGAGCAGTCT 3' and 5'AAAGGATCCGCGGCGGAGG 3' for R, 5'TGTAAGAACGGGAGCAGTCT 3' and 5'AAAGGATCCGCGGCGGAGG 3' and for R, 5'TGTAAGAACGGGAGCAGTCT 3' and 5'AAAGGATCCGCGGCGGAGG 3'. The amplified competitor DNAs were isolated from 6% Nusieve agarose gels by electroelution onto dialysis membranes.

RESULTS

Detection of minisatellite specific DNA-binding proteins

Mammalian testis nuclear extracts were initially screened by southwestern analysis, using 32P-labelled double-stranded cloned human minisatellites 33.15, λMS1 and λMS31 (4,7). Numerous nuclear proteins showed DNA-binding activity, although most also formed stable complexes with non-minisatellite control DNAs (SV-CAT plasmid DNA [23], and bacteriophage lambda DNA [24]; data not shown). Use of a synthetic double-stranded minisatellite B containing 5 repeats of the core GTGGGCAGGAAG (Fig.1A) considerably reduced the background of aspecific DNA-binding proteins, to reveal a single major 40kd DNA-binding protein in nuclear extracts of mouse brain (Fig.1B). In nuclear extracts made from rabbit (Oryctolagus cuniculus), Xenopus laevis and Drosophila melanogaster, similar DNA-binding proteins were observed (40kd in rabbit, 35 and 39kd in Xenopus, and 37kd in Drosophila). In contrast, double-stranded oligonucleotides R, and R, which contain randomized versions of the minisatellite sequence of B, (Fig.1A), were unable to form stable complexes with any of these proteins (Fig.1B). This demonstrates the existence of a ubiquitous minisatellite-specific DNA-binding protein which we term Msbp-1. A, repeat unit sequences of synthetic minisatellites An, Rn aligned with the minisatellite core sequence. Deviations from the core are indicated by underlined lower case. B, DNA-binding proteins detected by An-Rn in southwester analysis of nuclear extracts from mouse brain.

The tissue distribution of Msbp-1

Msbp-1 (40kd) was detected in all mouse tissues tested (brain, heart, kidney, liver and testis nuclear extracts; Fig.1C, plus lung and pancreas; not shown). Variable levels of a second minisatellite-specific binding protein of 36kd were present in samples from most tissues. The amount of this 36kd protein relative to Msbp-1 also varied 2–3 fold between separate batches of liver nuclear extract (not shown), suggesting that the 36kd protein might be a degradation product of Msbp-1. A third minisatellite-specific DNA-binding protein of 61kd was detectable only in testis nuclear extracts.

Sequence specificity of Msbp-1 binding

To investigate the sequence specificity of the binding of Msbp-1 to minisatellite DNA, the ability of the mouse protein to bind to a series of synthetic 'minisatellites' was tested. In the first group of such molecules, An-Rn, the repeat units were varied in terms of sequence similarity and repeat length with respect to the core sequence GTGGGCAGGAAG in probe B (Fig.2A). As
Figure 3. Effects of repeat unit copy number on the efficiency of minisatellite binding to Msbp-1. Top, detection of DNA-binding protein in southwestern analysis of mouse brain nuclear extracts, using the double-stranded probes B₁⁻⁻⁻⁻⁻⁻ B₅ which contain 1-5 copies respectively of the core repeat (Fig.1A), in the absence (-) or presence of non-specific competitor DNA. Bottom, relative binding efficiency of B₁⁻⁻⁻⁻⁻⁻ to Msbp-1 estimated from scanning densitometry of a shorter exposure of the autoradiograph shown and corrected for the different molar specific activities of each double-stranded oligonucleotide.

Expected, B₅ with an identical repeat sequence to B₅ was bound by Msbp-1 (Fig.2B). D₅ with a two base substitution at the 3' end of the core, and A₅ corresponding to tandem repeats of the chi sequence, were also bound by Msbp-1 (Fig.2B). In contrast, the short repeat probes C₅ and E₅, and the highly divergent core repeat in F₅, did not form stable complexes with Msbp-1, even though poor binding to higher molecular weight proteins was observed. In the second group of synthetic minisatellites, the repeat units were identical to those of B₅, except that nucleotide transversions were systematically introduced along the length of the core sequence (for example in the first of the group, G₅, GTGGGCA... became TTGGGCA and so on; Fig.2A). All of these synthetic minisatellites were bound by Msbp-1 (Fig.2B). Thus Msbp-1 binding activity is disrupted only by major alteration to the core sequences, and possibly by shifts in the periodicity of the tandem-repeats as in C₅ and E₅.

Requirement of tandem-repeats for Msbp-1 binding
The effect of minisatellite repeats unit copy number on Msbp-1 binding was investigated using double-stranded oligonucleotide probes B₁⁻⁻⁻⁻⁻⁻ B₅ containing 1-5 repeat units respectively of the core sequence GTGGGCAGGAAG (Fig.3). Remarkably, Msbp-1 did not bind B₁ even in the absence of non-specific competitor DNA, and DNA-protein interactions were strongly dependant on repeat number, with a 10 fold increase in binding efficiency comparing B₅ with B₁ (Fig.3). We conclude that strong binding of DNA by Msbp-1 requires at least 60bp of minisatellite corresponding in this case to 5 repeat units of the core sequence GTGGGCAGGAAG.

Kinetic and competition studies on minisatellite-Msbp-1 interactions
The kinetics of binding of synthetic minisatellite B₅ by Msbp-1 was characterised by increasing the time of DNA-binding reactions (Fig.4A). With the normal binding reaction time of 2h the amount of minisatellite bound was not maximal, and it was found that after 16h, binding of the labelled probe approached saturation. As true competition can only be achieved once the conditions of binding allow saturation of the available Msbp-1 binding sites, a reaction time of 16h was used in subsequent competition experiments.

To further test the specficity of interactions between Msbp-1 and minisatellite DNA, increasing amounts of unlabelled oligonucleotide competitor was included in southwestern binding reactions and any inhibition of binding of labelled B₅ noted (Fig.4B). Competitor oligonucleotide B₅ efficiently competed for the binding of labelled B₅ by Msbp-1, indicating that the binding
of B<sub>2</sub> to Msbp-1 is largely saturated at the concentration of probe used (3.3ng/ml). This implies a high affinity of Msbp-1 for oligonucleotide B<sub>2</sub> (K<sub>diss</sub> < 5 x 10<sup>-11</sup> M). In contrast, the randomized control oligonucleotides R<sub>1</sub> and R<sub>2</sub> showed far less efficient competition for binding of B<sub>2</sub>, providing further evidence for the sequence specificity of binding by Msbp-1. However both R<sub>1</sub> and R<sub>2</sub> showed some weak competitive activity, despite the fact that neither of these oligonucleotides can detect Msbp-1 by southwestern analysis.

**DISCUSSION**

Although the southwestern method for the detection of DNA-binding proteins has certain limitations, such as the poor blotting transfer of large proteins, and the inactivity of heteropolymers (18), the use of this method has allowed, the first detection of a minisatellite-specific DNA-binding protein to be made. This protein has an observed molecular weight of 40kd in the mouse and equivalent proteins have been detected in a number of other species including *Drosophila*. We cannot yet exclude the possibility that the observed Msbp-1 is a partial degradation product of a larger protein, which is either poorly transferred, or does not refold efficiently. Both murine Msbp-1 (40kd) and its presumed minisatellite-binding homologues in rabbit, Xenopus and *Drosophila* are unable to bind the randomized probes R<sub>1</sub> and R<sub>2</sub> (Fig.1A, 1B). Therefore binding by these proteins cannot simply depend on the presence of a GC-rich DNA sequence. Murine Msbp-1 shows high affinity (K<sub>diss</sub> < 5 x 10<sup>-11</sup> M), and is able to bind to a variety of synthetic minisatellites (Fig.2), indicating that binding is not absolutely dependent on perfect copies of the minisatellite core sequence. We note there is a variety in the sequences of cloned minisatellite loci (see e.g 2, 4, and 7) and it is tempting to speculate that the low sequence specificity of Msbp-1 binding may be related to this observed variation.

For strong interaction with DNA, Msbp-1 requires at least 60bp of minisatellite sequence, contained within the 91bp B<sub>2</sub> molecule (Fig.3). This is in contrast to other DNA-binding proteins which generally require <20bp for sequence specific binding of DNA (25). This might suggest that the complex between Msbp-1 and minisatellite B<sub>2</sub> involves binding of the DNA at a number of distinct sites within the 60bp of bound minisatellite. Thus with the proposed binding by Msbp-1 molecule(s) over this extended region, there is potential for changes to occur both in the tertiary structure and topology of the bound DNA. At present it is not clear whether efficient binding of Msbp-1 requires continuous tandem repetitions of the core sequence in the target DNA, or alternatively, whether binding is directed to specific and appropriately spaced copies of the core sequence present in the target.

The low inhibitory action of oligonucleotides R<sub>1</sub> and R<sub>2</sub> detected in the competition experiments (Fig.4B) suggests that stable binding of DNA by Msbp-1 is at least a two step process. Simplistically, this might be thought of as the binding of first one, then of a second minisatellite repeat unit by Msbp-1. If so, then GC-rich core-like sequences contained within R<sub>1</sub> and R<sub>2</sub> might possess some ability to inhibit the first step of the binding event. At present such ideas are speculative and await the results of further investigation.

The cellular function of Msbp-1 is not known, neither is any role, direct or indirect, that this protein may play in minisatellite mutation events. To further investigate its role it will be necessary to purify this protein. In an attempt to clone the cDNA for Msbp-1, we have screened a mouse brain λgt11 cDNA expression library (Clontech) with the double-stranded minisatellite probe B<sub>2</sub>, under conditions that should allow detection of any Msbp-1-lacZ fusion proteins possessing minisatellite-binding activity (26, 27). However, after screening 2.0 x 10⁶ plaques we have not been able to isolate any recombinant phage which contain such activity. We are therefore beginning the biochemical purification of this novel DNA-binding protein, Msbp-1.

**ACKNOWLEDGEMENTS**

We thank Martin Couchman for technical assistance. Also we thank John Keyte and Alec Markham for oligonucleotide synthesis, Nicola Royle for the gift of human and rat nuclear extracts, Ian Riddell for photography and Sarah Laband for typing.

This work was supported by a grants to A.J.J. from the Medical Research Council and the Wolfson Foundation. A.J.J. is a Lister Institute Research Fellow. The minisatellite probes are the subject of patent applications. Commercial enquiries should be addressed to Cellmark Diagnostics, 8 Blacklands Way, Abingdon Business Park, Abingdon, Oxon OX14 1DY, UK.

**REFERENCES**
